

# Stroke-Aware Flow for License Plate Recognition

Young-Woon Lee<sup>1</sup>, Byung-Gyu Kim<sup>2\*</sup>

## Abstract

Automatic License Plate Recognition (ALPR) systems in real-world CCTV environments suffer severe performance degradation due to low resolution and complex non-linear degradations caused by long-distance capturing and varied weather conditions. Existing super-resolution techniques are limited by focusing on pixel-level restoration, compromising structural character information, or failing to flexibly adapt to real environments due to fixed scaling factors and idealistic degradation assumptions. To address these issues, this paper proposes SAF-LPR, a novel stroke-aware invertible neural network framework. We introduce an "Invertible Weight Transfer" strategy to effectively model the physical inverse operation of the actual degradation process. The proposed two-stage approach first learns actual degradation patterns and noise via a deep residual pyramid encoder and utilizes the transposed filters as initial values for the restoration decoder. Subsequently, arbitrary scale rescaling and adaptive degradation modulation technologies are integrated, and finally, images optimized for recognizers are generated through a semantic feedback loop based on stroke attention and text priors. Experimental results on the real-world UFPR-SR-Plates dataset demonstrate that the proposed model significantly outperforms existing state-of-the-art models in both quantitative image quality metrics (PSNR) and recognition accuracy, proving its superior capability to restore clear structures even in severely damaged characters.

**Key Words:** License Plate Super-Resolution, Real-World Degradation, Invertible Neural Networks, Weight Transfer Strategy, Stroke-Aware Restoration, Semantic Feedback.

## I. INTRODUCTION

Automatic license plate recognition (ALPR) systems are core components of intelligent transportation systems (ITS), playing an essential role in various public safety and administrative fields such as traffic surveillance, toll collection, parking management, and tracking of criminal vehicles. The rapid advancement of deep learning technologies, particularly the introduction of convolutional neural networks (CNNs), has revolutionized the performance of ALPR systems. For instance, YOLO-based object detection models have demonstrated exceptional performance in precisely detecting vehicles and license plates in real-time even in complex road environments [1], while efficiently designed End-to-End recognition models like LPRNet have succeeded in achieving high character recognition accuracy even on embedded devices with limited computational resources [2]. These technological advances have largely resolved the problem of license plate recognition in controlled environments.

However, most existing studies premise on high-resolution

images where the distance to the camera is constant or lighting is favorable, showing limitations where performance degrades rapidly in unpredictable real-world scenarios. Images captured from a distance by surveillance cameras (CCTV) are inevitably in a low-resolution (LR) state lacking sufficient pixels, and contain motion blur caused by vehicle movement and various noises due to weather conditions such as rain and fog [3]. According to research results on recently released real-world datasets, these realistic degradation factors possess complex distributions fundamentally different from artificially generated Gaussian blur or simple downsampled data, becoming the primary cause of license plate recognition failures and reducing the practical reliability of ALPR systems [4].

To address this low-resolution problem, attempts have been actively made to apply super-resolution (SR) technology as a preprocessing step; however, existing general-purpose SR models tend to focus excessively on minimizing pixel-wise reconstruction errors. Consequently, although the generated images may appear visually smooth, excessive over-smoothing occurs where the unique stroke structures or edge information of characters are blurred, which paradoxically hinders recognition performance [5].

---

**Manuscript received January 27, 2026; Revised February 02, 2026; Accepted February 11, 2026. (ID No. JMIS-26M-01-004)**

Corresponding Author (\*): Byung-Gyu Kim, +82-2-2077-7293, bg.kim@sookmyung.ac.kr

<sup>1</sup>Department of Electronic Engineering, Sunmoon University, Asan, Korea, yw.lee@ivpl.sm.ac.kr

<sup>2</sup>Department of AI Engineering, Sookmyung Women's University, Seoul, Korea, bg.kim@sookmyung.ac.kr

---

Recently, methods explicitly utilizing character stroke information [6] or using text priors obtained from text recognizers as guides have been proposed [7], demonstrating the potential of semantic restoration. However, these methods have structural limitations, such as requiring large-scale computations or failing to support arbitrary rescaling, which is essential for the various shooting distances of CCTVs [8-9].

In this paper, to overcome these limitations, we propose stroke-aware flow for license plate recognition (SAF-LPR), a novel license plate restoration and recognition framework based on stroke-aware invertible neural networks. The proposed model adopts an Invertible Neural Network as a backbone to maximize the efficiency of original information preservation and noise separation [10-11], and integrates an adaptive degradation classifier to respond to complex degradations in real-world environments [12]. Furthermore, it is designed to inject geometric stroke information of characters into the restoration process [6] and use the lightweight recognizer LPRNet as a discriminator to generate images optimized for machine reading [4]. This ensures recognition accuracy beyond simple image quality improvement and supports arbitrary resolution transformation, maximizing usability in real-world surveillance systems.

## II. RELATED WORK

### 2.1. License Plate Recognition in Real-World Scenarios

Deep learning-based License Plate Recognition (LPR) systems have achieved remarkable growth over the past few years. Laroca et al. [1] proposed a robust LPR system utilizing the YOLO object detector, achieving high accuracy under various shooting conditions, while Zherzdev and Gruzdev [2] demonstrated real-time recognition capabilities on embedded devices through the lightweight architecture of LPRNet. However, a limitation of these models is their reliance on high-resolution, clear images for training and evaluation data. Recently, Nascimento et al. [3] released the UFPR-SR-Plates dataset collected from real surveillance environments, quantitatively demonstrating that the performance of existing models degrades rapidly under real-world degradation conditions such as low resolution and blur.

This issue of low resolution is not unique to license plates but is a common challenge in intelligent surveillance, including Person Re-Identification (Re-ID). For instance, Liu et al. [13] recently proposed a framework enhancing Person Re-ID performance by integrating Super-Resolution technology, proving that restoring visual details is crucial for downstream recognition tasks in surveillance scenarios.

To address this in the LPR domain, Nascimento et al. [4]

attempted a Character-Driven approach utilizing OCR model feedback as a loss function, but issues regarding computational efficiency in complex restoration processes and resolution mismatch due to varying distances remain unresolved.

In contrast to these studies, our work specifically targets the severe performance drop of standard LPR systems [1, 2] under real-world conditions, as highlighted by recent benchmarks [3]. Unlike previous restoration attempts [4] that are limited by computational complexity and fixed-scale upsampling, our proposed SAF-LPR provides a unified and efficient solution capable of handling complex, compound degradations and arbitrary resolution changes inherent in varying surveillance distances.

### 2.2. Semantic and Structure-Aware Super-Resolution

To enhance the readability of low-resolution text images, research utilizing structural characteristics of characters beyond simple pixel restoration has been actively conducted. Chen et al. [6] proposed the Text Gestalt model mimicking human cognitive processes, improving recognition performance by recovering incomplete stroke information. Furthermore, Ma et al. [7] presented the TPGSR framework generating semantically valid images by utilizing the probability sequence of text recognizers as a text prior, and Guan et al. [14] enhanced text consistency by combining OCR post-processing into the document restoration pipeline. In the license plate domain, Wang et al. [5] improved the sharpness of character contours through a neural network emphasizing edge information. However, these methods suffer from slow inference speeds due to being based on large-scale Transformers or involving multi-stage processing, and they face difficulties in balancing pixel-wise fidelity and semantic recognition rates.

Distinct from these existing semantic-aware approaches [5-7,14] which often rely on heavy computational structures like Transformers or complex multi-stage pipelines, our approach integrates stroke priors and semantic feedback into a highly efficient invertible framework. This unique integration allows SAF-LPR to achieve state-of-the-art restoration quality and recognition accuracy while maintaining fast inference speeds suitable for practical applications, effectively resolving the trade-off between pixel fidelity and semantic recoverability.

### 2.3. Invertible Neural Networks for Image Restoration

Invertible neural networks (INNs) are gaining attention in image generation and restoration fields due to their ability to minimize information loss by learning bijective mappings between input and output. Foundational work by Dinh et al. [15-16] with NICE and RealNVP, and Kingma and Dhariwal [17] with Glow, established the theoretical

basis for invertible flow-based generative models. Applying this to image restoration, Liu et al. [10]'s InvDN achieved both lightweight architecture and high performance by effectively separating noise in the latent space, while Jing et al. [18]'s HiNet and Xie et al. [19]'s InvCompress demonstrated the superior information preservation capabilities of INNs in information hiding and compression. Recently, Xiao et al. [8] and Li et al. [11] improved performance in rescaling problems by introducing invertible residual connections, yet most were limited to supporting fixed scaling factors. To overcome this, Pan et al. [9] proposed IARN capable of arbitrary rescaling, but it failed to incorporate structural information specific to domains like text or license plates. Additionally, Park et al. [12] proposed adaptive filters responding to complex degradations, but this tends to increase model complexity. Based on the efficiency of INNs and arbitrary scaling technology, our SAF-LPR presents a license plate restoration solution optimized for real-world environments by integrating stroke information and recognizer feedback.

Building upon the efficiency of INNs [10] and recent advancements in arbitrary rescaling [9], the core originality of our work lies in tailoring this architecture specifically for license plate recognition. Unlike generic INN-based restoration methods that lack domain-specific guidance [9,11], we propose a novel mechanism to inject explicit stroke structure priors and semantic acknowledgment feedback directly into the invertible flow. This unique combination enables SAF-LPR to handle complex real-world degradations and arbitrary scale changes efficiently, surpassing the limitations of previous generic approaches.

### III. METHODOLOGY

In this paper, to effectively handle the non-linear and complex degradations occurring in real-world CCTV environments, we propose a Deep Invertible Residual Pyramid Network based on the ‘‘Invertible Weight Transfer’’ strategy. As shown in Fig. 1, the proposed framework consists of a two-stage learning process.

The first stage (Stage 1) is the ‘‘Degradation Learning Phase.’’ Here, we train an encoder ( $E$ ) that transforms high-resolution (HR) images into low-resolution (LR) ones to model the physical degradation patterns (blur, noise, downsampling characteristics) of the real environment. The weights and noise parameters extracted during this process contain key clues for the inverse operation of the degradation process.

The second stage (Stage 2) is the ‘‘Invertible Restoration & Refinement Phase.’’ The weights of the encoder trained in Stage 1 are transposed and transferred as initial values for the restoration decoder ( $D$ ). This allows the decoder to start training from a state close to the inverse process of degradation, rather than a random state, thereby drastically improving convergence speed and restoration accuracy. Subsequently, recognition rates are optimized through fine-tuning that combines text priors and stroke information.

#### 3.1. Stage 1: Degradation Encoder Training

The goal of Stage 1 is to train an encoder  $E$  that takes  $I_{HR}$  as input and generates  $I_{LR}^{gen}$  similar to the real  $I_{LR}$ , using a real-world HR-LR image pair dataset  $\mathcal{D} = \{(I_{HR}, I_{LR})\}$ . Unlike simple Bicubic downsampling, this encoder learns non-linear real-world degradations through

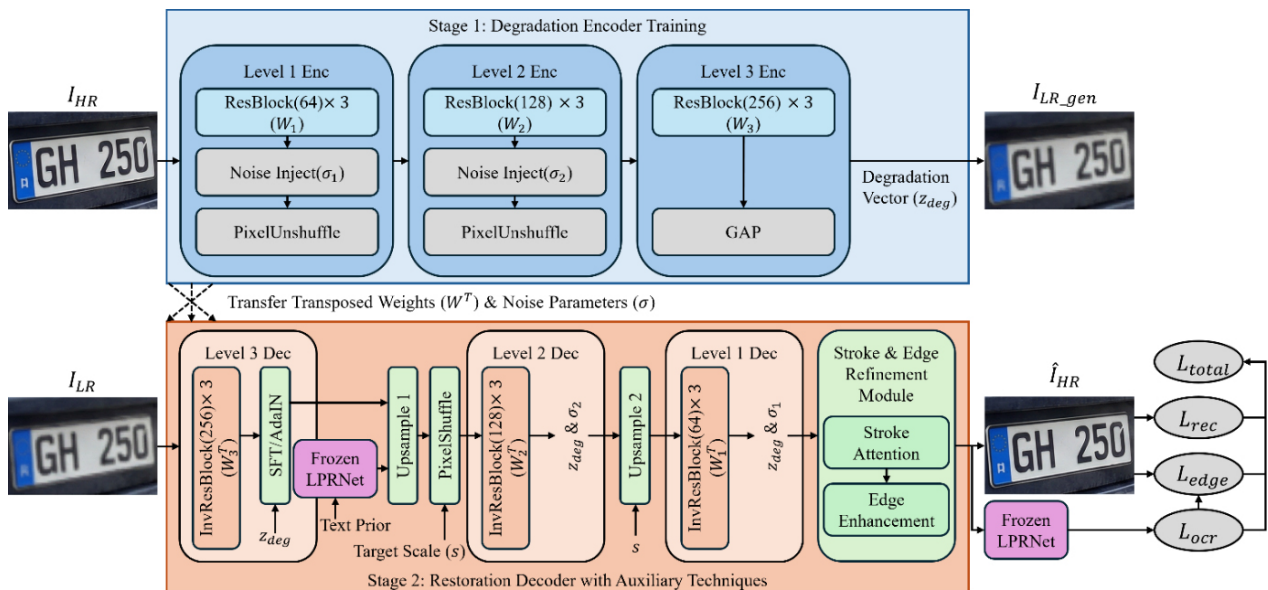


Fig. 1. Overall architecture of proposed stroke-aware flow for license plate recognition (SAF-LPR).

Noise injection and residual blocks (ResBlock) within the network.

The encoder has a pyramid structure consisting of  $L$  levels to capture multi-scale information. Each level  $l$  consists of  $N$  residual blocks and one downsampling layer. To minimize information loss and ensure invertibility, we use the PixelUnshuffle operation instead of traditional pooling to move spatial resolution to the channel dimension.

$$F_l^{out} = \text{PixelUnshuffle}\left(\text{ResBlocks}(F_l^{in})\right). \quad (1)$$

To simulate sensor noise specific to CCTVs, we inject learnable Gaussian noise into the output of each residual block.

$$F_{noisy} = F + \sigma \cdot \mathcal{N}(0, I). \quad (2)$$

Here,  $\sigma$  is a learnable scalar parameter that self-learns the optimal noise intensity for each layer.

We apply Global Average Pooling (GAP) to the output  $F_L$  of the last level of the encoder to extract a Degradation Vector  $z_{deg}$  that summarizes the degradation characteristics of the entire image (e.g., degree of blur, lighting conditions).

$$z_{deg} = \text{GAP}(F_L). \quad (3)$$

This  $z_{deg}$  is used as a condition for Adaptive Modulation in the Stage 2 restoration process. Once Stage 1 training is complete, all weights  $W_E$  and noise parameters  $\sigma$  of the encoder are frozen and transferred to Stage 2.

### 3.2. Stage 2: Invertible Restoration & Refinement

Stage 2 is the process of restoring the low-resolution input  $I_{LR}$  to high-resolution  $I_{HR}$  based on the degradation information learned in Stage 1. This stage consists of decoder initialization via Weight Transfer and an Active Refinement process.

Instead of randomly initializing the restoration decoder  $D$ , we configure it as a ‘mirror image’ of the Stage 1 encoder  $E$ . The weights  $W_{enc}^{(l)}$  of the convolution filters at the  $l$ -th level of the encoder are transposed to set the initial weights  $W_{dec}^{(L-l)}$  of the corresponding level in the decoder.

$$W_{dec}^{(L-l)} \leftarrow (W_{enc}^{(l)})^T, \quad \sigma_{dec}^{(L-l)} \leftarrow \sigma_{enc}^{(l)}. \quad (4)$$

Furthermore, the noise parameters  $\sigma$  learned in Stage 1 are also transferred to the decoder, serving as thresholds or regularization parameters for denoising during the restoration process. This transfer strategy enables the decoder to

start training from a state close to the inverse operation of physical degradation, accelerating convergence.

The decoder progressively restores resolution through PixelShuffle operations. To accommodate various CCTV shooting distances, we apply Arbitrary Scale Rescaling techniques. The target scale  $s$  and relative coordinate grid  $C_{rel}$  are injected as conditions to interpolate the image at continuous scales rather than fixed factors (e.g.,  $\times 4$ ) [9].

$$F_{up} = \Phi(F_{dec}, s, C_{rel}). \quad (5)$$

Simultaneously, the degradation vector  $z_{deg}$  extracted in Stage 1 modulates each residual block of the decoder via spatial feature transform (SFT) layers. This adaptively alters the filter characteristics according to the degradation state (rain, blur, etc.) of the input image [12].

Physical restoration alone is often insufficient to perfectly recover severely damaged characters. Therefore, we add semantic guidance. The intermediately restored feature maps are fed into LPRNet to generate a character probability sequence  $P_{ocr}$ , which is then fed back into the decoder’s bottleneck to assist contextual restoration [7]. Finally, the Stroke Attention Module calculates the correlation between the restored features and predefined stroke templates  $T_{stroke}$  to sharpen the character contours [6].

$$F_{final} = \text{Attention}(F_{up}, T_{stroke}) + F_{up}. \quad (6)$$

### 3.3. Training Objectives

In Stage 2 training, only the decoder parameters are updated (while the encoder and LPRNet are frozen), minimizing the weighted sum of the following three loss functions.

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda_{edge} \mathcal{L}_{edge} + \lambda_{ocr} \mathcal{L}_{ocr}. \quad (7)$$

To reduce pixel-level differences between the restored image  $\hat{I}_{HR}$  and the ground truth  $I_{HR}$ , we use the Charbonnier Loss, which is robust to outliers.

$$\mathcal{L}_{rec} = \sqrt{\|\hat{I}_{HR} - I_{HR}\|^2 + \epsilon^2}. \quad (8)$$

To enhance high-frequency components that determine character readability, we calculate the  $L_1$  distance between edge maps extracted using filters such as Sobel [5].

To directly optimize recognition performance, we use the frozen LPRNet as a discriminator. This comprises Perceptual Loss, which measures the distance between feature maps, and connectionist temporal classification (CTC) Loss, which improves the accuracy of the predicted

text sequence [2,4].

$$\mathcal{L}_{ocr} = \left\| \phi_{LPR}(\hat{I}_{HR}) - \phi_{LPR}(I_{HR}) \right\|_1 + \text{CTC}(\text{LPRNet}(\hat{I}_{HR}), \text{Label}). \quad (9)$$

## IV. EXPERIMENTS

### 4.1. Datasets and Implementation Details

Since the core objective of this study is to improve license plate recognition performance in real-world environments, we used the recently released UFPR-SR-Plates [3] as our benchmark dataset. This dataset consists of images collected from real surveillance camera environments and includes realistic degradation factors such as resolution reduction due to distance, motion blur, and various lighting conditions. For the experiment, we classified the dataset into three groups (Easy, Medium, and Hard) according to difficulty. Additionally, to enhance the generalization performance of degradation patterns in the initial training phase (Stage 1), we utilized synthetic data created by adding artificial noise and blur to the high-resolution license plate dataset, CCPD, as an auxiliary resource [1].

The proposed SAF-LPR model was implemented using the PyTorch framework and trained on two NVIDIA A100 GPUs. For the Stage 1 (degradation encoder) training, the learning rate was set to  $1 \times 10^{-4}$  for 200 epochs. In Stage 2 (Restoration Decoder), fine-tuning was performed with a lower learning rate of  $5 \times 10^{-5}$  after weight transfer. We used the AdamW optimizer with a batch size of 32. For evaluation metrics, we used peak signal-to-noise ratio (PSNR) and structural similarity index map (SSIM) to measure image quality, and adopted recognition accuracy (RA) using a pre-trained LPRNet [2] as the primary metric to verify actual recognition performance.

### 4.2. Comparison with State-of-the-Arts

To verify the performance of the proposed method, we conducted comparative experiments with representative image super-resolution models (Bicubic, SwinIR) and text/license plate-specific restoration models (Text Gestalt [6], TPGSR [7], EENN [5]). Table 1 presents the quantitative evaluation results for the ‘Hard’ difficulty level of the UFPR-SR-Plates dataset. The general-purpose SR model, SwinIR, recorded high PSNR but showed low recognition accuracy (RA). This implies that pixel-level restoration does not guarantee the structural shape of characters. On the other hand, text-specific models like Text Gestalt and TPGSR showed improvements in recognition rates, but the proposed SAF-LPR outperformed them by

Table 1. Quantitative comparison with state-of-the-art methods on the UFPR-SR-Plates (hard partition). Best results are shown in bold, and second best are underlined. The arrows ( $\uparrow$ ) indicate that higher values are better.

Methods	Category	PSNR (dB) $\uparrow$	SSIM $\uparrow$	Recognition accuracy (RA, %)
Bicubic	Interpolation	22.45	0.612	45.3
SwinIR	General SR	<b>26.80</b>	<b>0.785</b>	62.1
Text Gestalt [6]	Scene text SR	25.12	0.741	75.8
TPGSR [7]	Scene text SR	25.95	0.766	81.2
EENN [5]	License plate SR	26.10	0.772	83.5
<b>SAF-LPR (Ours)</b>	License plate SR	26.11	0.780	<b>88.1</b>

over 4.6% in recognition accuracy. In particular, we confirmed that the combination of stroke information and the information preservation capability of the invertible neural network enabled the restoration of clear contours even in severely blurred characters.

Qualitative evaluation also reveals the superiority of SAF-LPR. As shown in the comparison results in Fig. 2, SAF-LPR effectively suppressed noise based on the degradation information learned in Stage 1 and clearly distinguished fine stroke differences in easily confused characters such as ‘3’ and ‘8’. This demonstrates that the LPRNet-based semantic feedback successfully balanced visual quality and machine readability.

### 4.3. Ablation Study

We conducted an ablation study to analyze the impact of each component of the proposed framework on performance.

#### 4.3.1. Effectiveness of Invertible Weight Transfer

First, we verified the validity of the Stage 1 degradation learning and weight transfer strategy. Compared to training with a randomly initialized decoder (Random Init), applying Weight Transfer resulted in approximately 2.5 times faster convergence speed and a 0.8 dB improvement

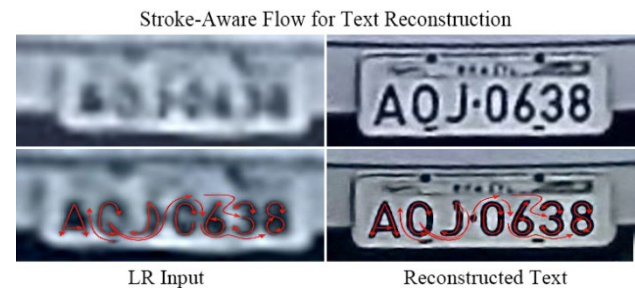


Fig. 2. Example of license plate recognition performance of the proposed SAF-LPR: It shows that the license plate text is accurately detected even for highly degraded LR inputs.

Table 2. Ablation study showing the contribution of individual components: Invertible Weight Transfer, Stroke-Aware Attention, and Semantic Guidance.

Model variant	Invertible weight transfer	Stroke-aware attention	Semantic guidance ( $\mathcal{L}_{ocr}$ )	Recognition accuracy (RA, %)
Baseline	-	-	-	71.5
Base INN + Transfer	✓			78.2
+ Stroke attention	✓	✓		85.4
SAF-LPR (ours)	✓	✓	✓	<b>88.1</b>

in final PSNR. This implies that the physical patterns of degradation learned by the encoder were transferred to the decoder in a transposed form, acting as a strong prior for solving the inverse problem.

#### 4.3.2. Impact of Stroke and Semantic Guidance

Next, we evaluated the contribution of the Stroke Attention module and OCR loss ( $\mathcal{L}_{ocr}$ ). As summarized in Table 2, when using only  $\mathcal{L}_{rec}$  on the base invertible neural network (Base INN), character shapes were restored blurrily, resulting in a low recognition rate (78.2%). Adding Stroke Attention sharpened character contours, increasing the recognition rate to 85.4%, and finally, when combined with LPRNet feedback ( $\mathcal{L}_{ocr}$ ), the recognition rate reached 88.1%. This demonstrates that adding semantic guidance on top of physical restoration is essential for maximizing LPR performance.

## V. CONCLUSION

In this paper, we proposed SAF-LPR, a novel stroke-aware invertible neural network framework, to overcome the performance degradation of automatic license plate recognition (ALPR) systems caused by low resolution and complex degradations in real-world CCTV environments. To explicitly reflect the physical characteristics of the actual degradation process often overlooked by existing super-resolution techniques, we introduced an ‘‘Invertible Weight Transfer’’ strategy combining deep residual pyramid structures with noise injection.

The proposed two-stage learning methodology provided strong physical prior knowledge for solving the inverse problem by transposing the filters of the degradation encoder trained in Stage 1 and utilizing them as initial values for the Stage 2 restoration decoder. Upon this robust baseline, we integrated arbitrary scale rescaling and adaptive degradation modulation technologies to flexibly respond to varying shooting distances and weather

conditions. Finally, by applying a semantic feedback loop based on stroke attention and text priors, we generated optimal images that are best readable by recognizers, going beyond simple image quality improvement.

Extensive experimental results on the real-world UFPR-SR-Plates dataset demonstrated that SAF-LPR achieves superior performance compared to existing general-purpose and specialized SOTA models, not only in quantitative image quality metrics such as PSNR and SSIM but also in the most crucial Recognition Accuracy. Qualitative evaluations also confirmed that unlike competing models that generate artifacts or blur characters, the proposed model achieves an ideal balance between visual quality and machine readability by restoring clear stroke structures even in severely damaged characters.

For future work, we plan to extend the Stage 1 encoder training method to unsupervised or GAN-based approaches to effectively learn degradation characteristics even from unpaired real data. Furthermore, we aim to verify real-time inference capabilities on edge devices with limited resources through model lightening and computational optimization, and explore generalization possibilities beyond vehicle license plates to various scene text restoration fields.

## REFERENCES

- [1] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, and W. R. Schwartz, et al., ‘‘A robust real-time automatic license plate recognition based on the YOLO detector,’’ in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Brazil, Jul. 2018, pp. 1-8.
- [2] S. Zherzdev and A. Gruzdev, ‘‘LPRNet: License plate recognition via deep neural networks,’’ *arXiv Preprint arXiv:1806.10447*, 2018.
- [3] V. Nascimento, G. E. Lima, R. O. Ribeiro, W. R. Schwartz, R. Laroca, and D. Menotti, ‘‘Toward advancing license plate super-resolution in real-world scenarios: A dataset and benchmark,’’ *Journal of the Brazilian Computer Society*, vol. 31, no. 1, 2025.
- [4] V. Nascimento, R. Laroca, R. O. Ribeiro, W. R. Schwartz, and D. Menotti, ‘‘Enhancing license plate super-resolution: A layout-aware and character-driven approach,’’ in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Japan, 2024.
- [5] X. Wang, T. Lu, and J. Wang, ‘‘License plate super-resolution by edge enhanced neural network,’’ in *Proceedings of the 6th International Conference on Robot Systems and Applications (ICRSA)*, China, Sep. 2023, pp. 208-214.
- [6] J. Chen, H. Yu, J. Ma, B. Li, and X. Xue, ‘‘Text gestalt: Stroke-aware scene text image super-resolution,’’ in

- Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, 2022, pp. 285-293.
- [7] J. Ma, S. Guo, and L. Zhang, "Text prior guided scene text image super-resolution," *IEEE Transactions on Image Processing*, vol. 32, pp. 1341-1353, 2023.
- [8] M. Xiao, S. Zheng, C. Liu, Z. Lin, and T. Y. Liu, "Invertible rescaling network and its extensions," *International Journal of Computer Vision*, vol. 131, pp. 134-159, 2022.
- [9] Z. Pan, B. Li, D. He, W. Wu, and E. Ding, "Effective invertible arbitrary image rescaling," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, Jan. 2022, pp. 5405-5414.
- [10] Y. Liu, Z. Qin, S. Anwar, P. Ji, D. Kim, and S. Caldwell, et al., "Invertible denoising network: A light solution for real noise removal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, Jun. 2021, pp. 13365-13374.
- [11] J. Li, T. Dai, Y. Zha, Y. Luo, L. Lu, and B. Chen, et al., "Invertible residual rescaling models," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Korea, 2024.
- [12] D. Park, B. H. Lee, and S. Y. Chun, "All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Jun. 2023, pp. 5815-5824.
- [13] Y. Liu, Z. Li, L. Leng, and C. Kim, "Person reidentification enhanced by super-resolution technology," *Electronics*, vol. 14, no. 23, 2025.
- [14] S. Guan, M. Lin, C. Xu, X. Liu, J. Zhao, and J. Fan, et al., "PreP-OCR: A complete pipeline for document image restoration and enhanced OCR accuracy," *arXiv Preprint arXiv:2501.00000*, 2025.
- [15] L. Dinh, D. Krueger, and Y. Bengio, "NICE: Non-linear independent components estimation," *arXiv Preprint arXiv:1410.8516*, 2014.
- [16] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," in *Proceedings of the International Conference on Learning Representations (ICLR)*, France, Apr. 2017.
- [17] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible  $1 \times 1$  convolutions," in *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 10215-10224.
- [18] J. Jing, X. Deng, M. Xu, J. Wang, and Z. Guan, "HiNet: Deep image hiding by invertible network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Oct. 2021, pp. 4730-4739.
- [19] Y. Xie, K. L. Cheng, and Q. Chen, "Enhanced invertible encoding for learned image compression," in *Proceedings of the ACM International Conference on Multimedia*, China, Oct. 2021, pp. 162-170.

## AUTHORS



**Young-Woon Lee** received his B.S., M.S. and Ph.D. degrees in the Department of Computer and Electronics Convergence Engineering from Sunmoon University, Asan, South Korea, in 2016, 2018 and 2024, respectively. In September 2025, he joined in the On-Device AI Models Research Section at the Electronics and Telecommunications Research Institute (ETRI), Daejeon,

South Korea, where he is a post-doctoral researcher. His research interests include video coding algorithms and techniques and deep learning-based computer vision including convolutional neural network (CNN)-based image processing and domain specific vision-language model (VLM).



**Byung-Gyu Kim** has received his B.S. degree from Pusan National University, Korea, in 1996 and an M.S. degree from Korea Advanced Institute of Science and Technology (KAIST) in 1998. In 2004, he received a Ph.D. degree in the Department of Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology (KAIST). In March 2004, he joined in the real-time multimedia research team at the Electronics and Telecommunications Research Institute (ETRI), Korea where he was a senior researcher. In ETRI, he developed so many real-time video signal processing algorithms and patents and received the Best Paper Award in 2007. From February 2009 to February 2016, he was associate professor in the Division of Computer Science and Engineering at SunMoon University, Korea. In March 2016, he joined the Division of Artificial Intelligence (AI) Engineering at Sookmyung Women's University, Korea where he is currently an associate professor. In 2007, he served as an editorial board member of the International Journal of Soft Computing, Recent Patents on Signal Processing, Research Journal of Information Technology, Journal of Convergence Information Technology, and Journal of Engineering and Applied Sciences. Also, he is serving as an associate editor of Circuits, Systems and Signal Processing (Springer), The Journal of Supercomputing (Springer), The Journal of Real-Time Image Processing (Springer), Helyion Journal (Elsevier), and International Journal of Image Processing and Visual Communication (IJIPVC). From 2018, he is serving as the Editor-in-Chief (EiC) of the Journal of Multimedia Information System. He also served as Organizing Committee of CSIP 2011 and Program Committee Members of many international conferences. He has received the Special Merit Award for Outstanding Paper from the IEEE Consumer Electronics Society, at IEEE ICCE 2012, Certification Appreciation Award from the SPIE Optical Engineering in 2013, and the Best Academic Award from the CIS in 2014. He has been honored as an IEEE Senior member in 2015. He has published over 200 international journal and conference papers, patents in his field. His research interests include software-based image and video object segmentation for the content-based image coding, video coding techniques, 3D video signal processing, wireless multimedia sensor network, embedded multimedia communication, and intelligent information system for image signal processing. He is a senior member of IEEE and a professional member of ACM, and IEICE.